Réseaux de gènes : inférence, évaluation, utilisation et au-delà

Nathalie Vialaneix

Réseaux de gènes : inférence, évaluation, utilisation et au-delà

Nathalie Vialaneix [1]

[1] *Université Fédérale de Toulouse, INRAE, MIAT, 31326 Castanet-Tolosan, France*

Résumé. La collecte de données à l'échelle moléculaire s'est considérablement accrue au cours des vingt dernières années, en quantité mais aussi en variété et précision. Cette évolution rapide crée un besoin de développement de méthodes d'analyse adaptées à la complexité et au volume de ces données : l'espoir pour les biologistes est que l'utilisation de ces nouvelles données ouvre la voie à une meilleure compréhension du fonctionnement du vivant et des relations complexes entre séquence d'ADN, environnement et ce que l'on peut observer à l'échelle de l'individu. Les répercussions potentielles sur le traitement des maladies (dont le cancer) ou la sélection des espèces agricoles animales et végétales pour faire face au changement climatique touchent à des questions sociétales importantes.

Une des données moléculaires les plus utilisées et étudiées pour caractériser le fonctionnement des cellules est l'*expression des gènes* (aussi appelée transcriptomique), qui est un mécanisme sous forte régulation génétique et épigénétique. Il est courant de représenter ces régulations sous la forme de graphes (ou réseaux) de gènes et la reconstruction de ces graphes, à partir de données d'expériences temporelles ou statiques, a été et demeure un sujet actif de recherche en statistique, connu sous le nom d'*inférence de réseaux de gènes* [11, 4, 7, 3, 10, 5, 6, 2, 8, 13]. À l'inverse, plusieurs méthodes de prédiction (régression ou classification) ont été développées pour inclure cette information de régulation sous forme de graphe et estimer à partir de celle-ci un phénotype mesuré à l'échelle de l'individu [12, 9].

Dans cet exposé, je dresserai un panorama des méthodes d'inférence de réseaux et de prédiction à base de graphes (en particulier des réseaux de neurones pour graphes [1]) et je discuterai les limites actuelles de leur utilisation ou de leur évaluation en regard de la complexité des mécanismes moléculaires modélisés.

Cette présentation inclut des résultats de travaux publiés ou en cours, réalisés en collaboration avec Céline Brouard, Anne Goelzer, Raphaël Mourad et Vincent Rocher.

Mots-clés. réseaux de gènes, transcriptomique, inférence de réseau, réseaux de neurones pour graphe

Abstract. Data collection at the molecular level has grown considerably during the last twenty years, not only in quantity but also in variety and and precision. This rapid evolution creates a need for the development of analysis methods adapted to the complexity and volume of these data. The hope for biologists is that the use of these new data will pave the way to a better understanding of how living organisms function and will unravel part of the complex relationships between DNA sequence, environment and what can be observed at the individual level. The potential repercussions include treatment of diseases (including

cancer) and the selection of agricultural plant and animal species able to cope with climate change, both being critical societal issues.

One of the most widely used and studied molecular data characterizing cell functioning is *gene expression* (also known as transcriptomics). Gene expression is a complex molecular mechanism that is a highly genetically and epigenetically regulated and it is a common practice to represent these regulations in the form of graphs (or networks) of genes. The reconstruction of these graphs using data from temporal or static experiments, has been and remains an active subject of statistical research, known as gene network inference [11, 4, 7, 3, 10, 5, 6, 2, 8, 13]. In addition, several prediction methods (regression or classification) have been developed to include these regulatory networks and to use them to better estimate a phenotype measured at the organism level [12, 9].

In this talk, I will give an overview of network inference and prediction methods based on graphs (and, in particular, on graph neural networks [1]). I will discuss the current limits of their use or evaluation with respect to the complexity of the molecular molecular mechanisms being modeled.

This presentation includes published and ongoing works made in collaboration with Céline Brouard, Anne Goelzer, Raphaël Mourad, and Vincent Rocher.

Keywords. gene networks, transcriptomics, network inference, graph neural networks

Bibliographie

[1] Céline Brouard, Raphaël Mourad, and Nathalie Vialaneix. Should we really use graph neural networks for transcriptomic prediction? *Briefings in Bioinformatics*, 25(2):bbae027, 2024.

[2] Océane Cassan, Sophie Lèbre, and Antoine Martin. Inferring and analyzing gene regulatory networks from multi-factorial expression data: a complete and interactive suite. *BMC Genomics*, 22:387, 2021.

[3] J. Chiquet, Y. Grandvalet, and C. Ambroise. Inferring multiple graphical structures. *Statistics and Computing*, 21(4):537–553, 2011.

[4] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[5] M. Gallopin, A. Rau, and F. Jaffrézic. A hierarchical Poisson log-normal model for network inference from RNA sequencing data. *PLoS ONE*, 8(10), 2013.

[6] Johann S. Hawe, Fabian J. Theis, and Matthias Heinig. Inferring interaction networks from multi-omics data. *Frontiers in Genetics*, 10:535, 2019.

[7] Vân Anh. Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE*, 5(9):e12776, 2010.

[8] Yoonjee Kang, Denis Thieffry, and Laura Cantini. Evaluating the reproducibility of single-cell gene regulatory network inference algorithms. *Frontiers in Genetics*, 12:362, 2021.

[9] Caiyan Li and Hongzhe Li. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24(9):1175–1182, 2008.

[10] Daniel Marbach, James C. Costello, Robert Küffner, Nicci Vega, Robert J. Prill, Diogo M. Camacho, Kyle R. Allison, the DREAM5 Consortium, Manolis Kellis, and Gustavo Collins, James J.and Stolovitsky. Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8):796–804, 2012.

[11] N. Meinshausen and P. Bühlmann. High dimensional graphs and variable selection with the Lasso. *Annals of Statistic*, 34(3):1436–1462, 2006.

[12] Franck Rapaport, Andrei Zinovyev, Marie Dutreix, Emmanuel Barillot, and Jean-Philippe Vert. Classification of microarray data using gene networks. *BMC Bioinformatics*, 8:35, 2007.

[13] Michael Saint-Antoine and Abhyudai Singh. Benchmarking gene regulatory network inference methods on simulated and experimental data. bioRxiv preprint, 2023.